科技论文初审查重结果判定模式研究

杨继涛[1)] 郭柏寿[2)*]

1） 西北农林科技大学经管学院，《陕西农业科学》编辑部，陕西杨凌邰城路 3 号，712100
2）西北农林科技大学农学院，《西北农业学报》编辑部，陕西杨凌邰城路 3 号，712100

摘要 【目的】旨在对科技论文初审查重结果判定模式进行研究，以期提出科学辩证的取舍策略，使论文初审更加公平与公正。【方法】首先采用调查统计法对当前初审查重现状进行分析，其次利用数学分析法对论文总复制比给予解构，最后根据相似片段分布对论文加以分类，针对不同类型提出不同的判定模式。【结果】调查统计发现，在目前各期刊的初审工作中，论文总复制比是否超过规定的复制比阈值已具备一票否决的作用，有失偏颇。本研究根据论文各部分的重要性以及相似片段在不同部分的分布情况，将初审论文分为连续分布型、一头分布型、两头分布型、中间分布型，其中连续分布型又分为连续均匀型和连续非均匀型；一头分布型再分为头部分布型和尾部分布型；在此基础上，对不同类型提出有针对性的判定策略，并给出论文查重辩证取舍实操工作流程与论文辩证查重取舍计算机程序框图。【结论】论文审稿工作是期刊出版工作的重要环节之一，不断探索优化审稿过程的每个流程尤为必要。随着计算机和信息技术的不断发展，基于辩证分析思想的论文人工智能查重系统一定会被开发出来，长期困扰期刊的学术质量把控及学术不端行为界定等问题也会得到更加合理优化地解决，将使期刊论文的审稿环节更加合理、客观与公正。

关键词 科技论文；学术不端；查重结果；复制比；公平公正

1 引言

自从有学术出版与学术交流以来，一稿多投、一稿多发、抄袭、剽窃等不端行为始终困扰着各出版社、期刊社及期刊编辑部。在网络与信息数字化技术未引入出版界之前，为杜绝学术不端现象，主要靠评审专家和编辑查阅相关领域的文献，并结合自身学科素养，通过比对分析进行把关，不但效率低下，而且也达不到预期效果。

近年来，网络和计算机信息技术在编辑出版领域得到广泛应用，有关机构也研发出基于互联网的学术不端文献检测系统，为用户提供文献查重服务。各期刊出版单位以此系统为依托，将相似度检测作为论文审核的必备事项[1]，相似度则由重复率（或称复制比）定量描述。时至今日，重复率检测（或文献查重）已成为如今学者发表期刊论文、科研项目结项及高校学生毕业论文答辩的第一步，因此顺利通过重复率检测至关重要[2]。

目前，国内相似度检测影响较大的有4个系统，分别属于中国知网、万方数据、Paperpass和维普，它们都根植于各自庞大的论文数据库，为用户提供对抄袭与剽窃、伪造、篡改等学术不端行为的快速检测功能，逐渐在会议论文、学位论文和期刊论文等的初审中发挥着越来越重要的作用。

几乎所有学术期刊都采用了独立的网上办公及采编系统，在线投稿是其主要途径甚至是大多数期刊的唯一投稿途径。尽管期刊采编系统种类较多，但基本都具备兼容学术不端检测模块的功能，因此当前各期刊普遍以采编系统嵌入的检测工具识别初审稿件的抄袭、剽窃等学术不端行为[3]。

早期的学术不端文献检测只限于高校对学位论文的评审上，后来在期刊论文的初审上也逐步引入学术不端检测。学术不端文献检测软件或为有关文献数据库的一部分，或嵌入期刊采编系统，如北京勤云科技发展有限公司的采编系统集成了万方数据的检测软件；西安三才公司的则集成了中国知网的检测软件；北京玛格泰克采编系统是国内最早投入使用的，历经

作者简介：杨继涛(ORCID: 0000 — 0003 — 2896 — 8614) ，硕士，副编审，《陕西农业科学》副主编，E-mail: snkx001@163.com
通讯作者郭柏寿(ORCID: 0000 — 0003 — 0119 — 2510) ，副编审，《西北农业学报》副主编，E-mail: xbnx04@ nwsuaf． edu． cn

十余载的改进和升级完善，目前技术较成熟、功能较完善、拥有较多用户，该采编系统设置了文献检索接口、元数据服务接口、参考文献检查与核对、转接审稿人智能推荐接口等，主要与国外 CrossRef 数据库合作，提供文献对比检测[4]。

从上述现状分析可知，各期刊社（编辑部）无论采用那种做法进行论文初审，在查重这块其判定初审论文是否涉嫌学术不端的依据均在于文献检测系统给出的总复制比是否超出期刊设定的复制比阈值。因为当一篇论文的总复制比很低时，基本可以认定其研究内容比较新颖、创新性较强、抄袭他人文献的可能性很低，应该不会涉及学术不端行为。由此可见，在当前各期刊初审工作中论文总复制比是否超过复制比阈值的裁判已具备一票否决的作用。

利用采编系统的检测工具获得总复制比，为期刊编辑在论文初审中判定学术不端行为提供了极大的方便。但随着这种模式在编辑实践中的普遍应用，其弊端和不足也渐渐显露出来，仅从期刊设定的总复制比阈值来判定论文的学术不端行为有时也显得草率和机械，常会造成误判，从而错失一些有创新性的稿件。因此对于论文初审中学术不端行为的判定有必要审慎对待，辩证地分析总复制比，并对总复制比进行解构分析，才能在初审环节体现出客观性与公正性。

2 概念辨析与研究方法

2.1 常规初审的判定模式

当前，期刊社或编辑部普遍的做法不外乎两种：（一）通过集成在采编系统里的文献检测软件先对新投稿件进行查重，若总复制比（以 r 表示，下同）超过期刊规定的上线（本文称之为复制比阈值）即做退稿处理，若总复制比不大于复制比阈值，则进入内容初审环节；（二）对论文先行内容初审（包括刊稿范畴、格式和学术价值等审查），不符合要求的做退稿处理，内容审查合格的再做初审的查重工作，若总复制比大于复制比阈值则给予退稿，若不大于复制比阈值即表示初审通过。

对复制比阈值的设定状况，笔者做了初步研究，调查对象为《中国学术期刊影响因子年报（自然科学与工程技术）2021 年版》收录的 103 种农业综合类期刊。调查结果表明，在所调查的期刊中，复制比阈值（r_0）的设置范围在 5%—35%之间。其中，r_0= 5% 的有 3 个期刊，5%<r_0≤10% 的有 8 个期刊，10%<r_0≤15% 有 14 个期刊，15%<r_0≤20% 的有 34 个期刊，20%<r_0≤25% 的有 25 个期刊，25%<r_0≤30% 的有 18 个期刊，30%<r_0≤35% 的有 1 个期刊。对所选期刊初审复制比阈值进行分析，具体结果如图 1 所示。

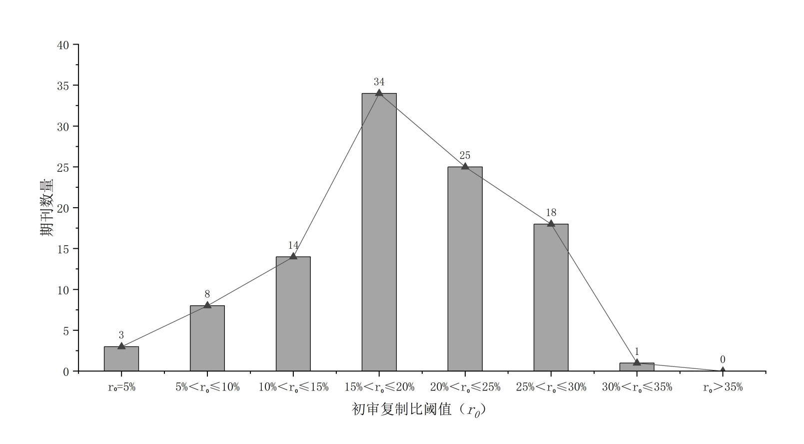


图 1　103 种农业综合类期刊复制比阈值（r_0）分布图

作者简介：杨继涛(ORCID: 0000 — 0003 — 2896 — 8614) ，硕士，副编审，《陕西农业科学》副主编，E-mail: snkx001@163.com

通讯作者郭柏寿(ORCID: 0000 — 0003 — 0119 — 2510) ，副编审，《西北农业学报》副主编，E-mail: xbnx04@ nwsuaf． edu． cn

2.2 总复制比解构

不论何种体裁或表现形式的论文，都可以分成相对独立的几节，这些节又可归于核心部或非核心部两类。当然核心部涉及的都是论文关键的、创新的、以及有别于其他论文的内容，非核心部涉及研究背景的综述以及与其他研究文献的比较分析等内容。由此可见，当核心部与非核心部复制内容的数量相同时，两者对被考察论文的负面影响可能是不同的，即使核心部的复制内容数量小于非核心部，这时对论文的伤害也许是致命的。因此，有必要对论文的总复制比进行结构，明确每节的复制比与总复制比的关系，结合其在总复制比中所占比率及该节对论文重要性的贡献大小，进行辩证考察与分析。

2.3 确定加权系数

为了论述方便，特以实验研究类论文为例进行总复制比结构。这类论文基本分为前言、材料与方法、结果与分析、讨论与结论及参考文献 5 节（摘要的内容主要抽提自“材料与方法”及“结果与分析”，可由二者替代，故略去），分别以 1、2、3、4、5 表示之。显然 2 和 4 可看作实验研究类论文的核心部，1、4 和 5 可视为非核心部。设 r_1'、r_2'、r_3'、r_4'和 r_5'为 1—5 节独自检测所得复制比，显然不能由 r_1'、r_2'、r_3'、r_4'和 r_5'直接相加而求总复制比 r 之值。必须对 r_1'—r_5'进行加权修正后，才能相加算出 r 值。以 N 代表论文的总字符数，n_1、n_2、n_3、n_4 和 n_5 各自代表 1—5 节的字符数，那么这 5 节的加权系数分别为：$\mu_1=n_1/N$，$\mu_2=n_2/N$，$\mu_3=n_3/N$，$\mu_4=n_4/N$，$\mu_5=n_5/N$；设加权修正后 1—5 节复制比依次为 r_1、r_2、r_3、r_4 和 r_5，那么可得：$r_1=\mu_1 r_1'=n_1 r_1'/N$，$r_2=\mu_2 r_2'=n_2 r_2'/N$，$r_3=\mu_3 r_3'=n_3 r_3'/N$，$r_4=\mu_4 r_4'=n_4 r_4'/N$，$r_5=\mu_5 r_5'=n_5 r_5'/N$；故可得：$r=r_1+r_2+r_3+r_4+r_5=n_1 r_1'/N+n_2 r_2'/N+n_3 r_3'/N+n_4 r_4'/N+n_5 r_5'/N=(n_1 r_1'+n_2 r_2'+n_3 r_3'+n_4 r_4'+n_5 r_5')/N$；若以 a_1、a_2、a_3、a_4 和 a_5 分别表示 1—5 节复制比占总复制比的比率，则知 $a_1=r_1/r$，$a_2=r_2/r$，$a_3=r_3/r$，$a_4=r_4/r$，$a_5=r_5/r$。接下来根据 a_1、a_2、a_3、a_4 和 a_5 的数值大小以及各节对论文的重要性，将论文分型，通过辩证分析，不做一刀切，对不同类型论文设置不同的复制比阈值，依此裁决，方可使初审工作更加公平与公正。

2.4 辩证分析论文分型

确定加权系数这一方法虽然是基于数量化分析的解构过程，但工作量比较大，需要分别将 5 节单独检测，还要统计论文的总字符数及各节的字符数。依据对总复制比的解构和加权分析，可以转换成更为便捷的论文分型方法，即依据论文中相似片段的分布进行分型来判定其初审中的学术不端行为。。

随机从采编系统中选取一篇实验研究类论文，从学术不端文献检测系统获得其查重所得的相似度分布图（图 2），从图 2 的相似片段分布可以看到，总复制比由论文头部、中前部、中部、中后部及尾部的复制内容字符数累加除以论文总字符数而得，这几部分大体对应实验研究类论文的 1—5 节。研判图 2 可直观地获悉每部分复制内容的多少及疑似学术不端的严重度（黄色表示“轻度抄袭”，橙色表示“中度抄袭”，红色表示“重度抄袭”， 绿色表示“没有抄袭”、即没有和其他文献重合的字数[7]），然后据此做论文分型。按分型类型对论文初审环节进行学术不端行为判定。

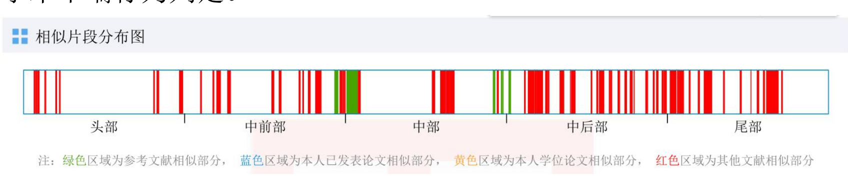


图 2 论文不同部分相似片段分布

作者简介：杨继涛(ORCID: 0000 － 0003 － 2896 － 8614) ，硕士，副编审，《陕西农业科学》副主编，E-mail: snkx001@163.com
通讯作者郭柏寿(ORCID: 0000 － 0003 － 0119 － 2510) ，副编审，《西北农业学报》副主编，E-mail: xbnx04@ nwsuaf． edu． cn

3 目前论文初审查重中存在的问题

仔细探究当下期刊在论文初审环节的普遍做法，可以发现，期刊编辑在行使初审权限时基本都是按采编系统提供的总复制比结合个刊设置的复制比阈值进行初审判定，并未对论文查重结果进行深入地辩证分析。这种做法存在的问题具体表现在：没有研究论文的总复制比是如何形成的，没有考虑论文各部分（前言、材料与方法、结果与分析、结论与讨论等）复制比与总复制比的关系如何，没有考察各部分的相似片段分布情况如何，没有意识到论文的哪些部分复制比过高才是真正的学术不端。可见，仅依据文献检测系统给出的总复制比是否超出期刊设置的唯一复制比阈值进行学术不端判定，不但简单轻率，而且有失偏颇。

对于一篇内容新颖的科技论文来说，其中存在一定比例的复制内容，并不意味着该论文彻底丧失科学价值，需要对论文中的复制内容进行溯源[5]。同时，论文中不同部分的文字重合通常所提示的学术不端性质也会存在明显差异[6]。

如前所述，科技论文的总复制比是由其各部分（或各章节）复制比加权累加形成的，因此，要做到公平公正，就需要对论文的总复制比进行细分，辩证分析论文不同构成部分各自复制比的大小及其对论文的负面影响程度，结合论文相似片段分布情况通盘考察其查重结果，使初审取舍工作尽量审慎客观。

4 解决方案

4.1 待审论文的分型及取舍原则

根据相似片段在相似度分布图中的分布情况，可把待审论文分为 4 种类型，即：连续分布型、一头分布型、两头分布型、中间分布型，其中连续分布型又可分为连续均匀型和连续非均匀型；一头分布型可再分为头部分布型和尾部分布型。搞清论文所属的类型后，就可以对症下药，具体问题具体分析，制定不同类型的取舍原则。

4.1.1 连续分布型及其取舍原则

图 3 和图 4 都属于连续分布型，这种类型的论文相似片段几乎分布于整篇论文的各个部分,也就是其全篇几乎都存在和其他文献相似的内容，明显没有多大学术价值，论文的创新性更无从谈起，建议做初审退稿处理。



图 3　论文相似片段的连续非均匀分布



图 4　论文相似片段的连续均匀分布

4.1.2 一头分布型及其取舍原则

从图 5 可以看出，一头分布型中的“头部分布类”的复制内容主要集中在论文的前言部分。前言一般都是对已发表的文献进行综合评述，阐明其研究的优点与不足，指出还有哪些问题需要继续探讨，并提出作者的研究目的与期许。当应用的文献发表时间较早，或许已被同领域或同专业其他论文反复应用，检测时可能被当做复制内容；另外，对于行业关注度较大的文献，作者引用时往往都以“直接引语”的形式加以表述，自然会出现复制比比较高的情况。

由图 6 可见，一头分布型中的“尾部分布类”，其相似片段主要分布在讨论与结论或参考文献部分，讨论部分的功能是：对前人研究结果与作者的结果进行比较分析，以说明作者的

作者简介：杨继涛(ORCID: 0000 － 0003 － 2896 － 8614) ，硕士，副编审，《陕西农业科学》副主编，E-mail: snkx001@163.com
通讯作者郭柏寿(ORCID: 0000 － 0003 － 0119 － 2510) ，副编审，《西北农业学报》副主编，E-mail: xbnx04@ nwsuaf． edu． cn

研究与前人在哪些方面相同或相似，又在哪些方面相左，并探讨结果相左产生的原因。因此，讨论部分和前言一样难免也要大量引用已报道的文献来作为论据，因此这部分出现较多的内容重复是可以理解的。论文结论部分的功能是：在结果与分析的基础上，提炼出具有普遍价值的科学规律。尾部分布型论文的结果与分析部分几乎没有重复内容，当具有较高的创新性，结论中不该出现高复制现象；反之，就是作者从实验结果未能发现真正规律性的东西，尚可修改凝练。若参考文献部分复制比高，极有可能是文献陈旧引起的，这对论文同样不足以造成致命性伤害，因为若以与作者研究相关的最新文献替代后，问题会得到解决。

可见，即使一头分布型论文的总复制比大于复制比阈值，也不足以否定论文在取材、研究方法、研究结果及结论上的客观性、合理性与创新性。在内容评审通过后的初审查重时，该型论文的复制比阈值可以适当定高一些，比如可令复制比阈值 r_0=25% 。当总复制比 r>25% 时，可给予初审退稿；r≤25% ，给予初审修改，可要求作者查阅最新文献，重新撰写前言或讨论与结论部分，以降低总复制比。



图 5 论文相似片段一头分布型中的头部类



图 6 论文相似片段一头分布型中的尾部类

4.1.3 两头分布型及其取舍原则

两头分布型（图 7）在初审论文中属于比较常见的类型，该型的相似片段主要集中于论文的前言、讨论与结论及参考文献部分，显而易见论文的核心部——材料与方法、结果与分析中没有和其他文献相重复的内容，若内容通过评审，尚有修改价值。鉴于前言、讨论与结论及参考文献合起来可能占到论文总篇幅的一半左右，对于两头分布型的复制比阈值可以定得比一头分布型更高些，比如令复制比阈值 r_0=35%。当总复制比 r>35% 时，建议初审退稿；r≤35% 则做初审退修处理，当然在初审退修意见中可要求作者认真修改，降低总复制比。

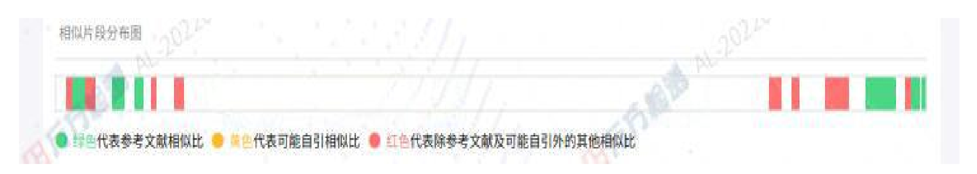


图 7 论文相似片段的两头分布型

4.1.4 中间分布型及其取舍原则

相似片段的中间分布型见图 8，如果再细分，还存在中前部、中部、中后部三种情况。中间分布型论文相似片段所处位置对应实验研究类论文的材料与方法及结果与分析两节，它们是论文的核心与灵魂。正因为如此，这型论文的复制比阈值应该定低一些，可令复制比阈值 r_0=5% ，若总复制比 r>5% ，修改的意义已经不大，可以考虑做初审退稿处理；若 r≤5%，尚可初审退修，但返修后最好再做一次查重。

作者简介：杨继涛(ORCID: 0000 － 0003 － 2896 － 8614) ，硕士，副编审，《陕西农业科学》副主编，E-mail: snkx001@163.com
通讯作者郭柏寿(ORCID: 0000 － 0003 － 0119 － 2510) ，副编审，《西北农业学报》副主编，E-mail: xbnx04@ nwsuaf． edu． cn

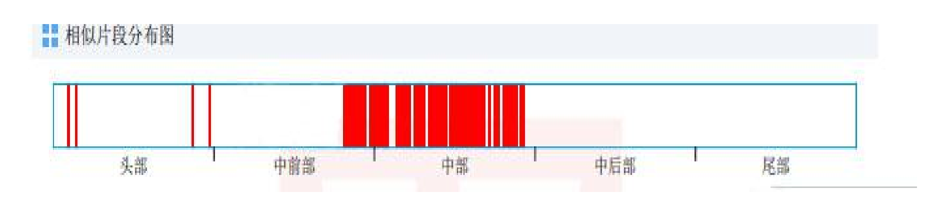


图 8　论文相似片段的中间分布型

4.2 论文初审查重辩证取舍实操流程

综合以上的辩证分析结果，编制了规范化的论文初审工作流程图（图 9），图中一头分布型、两头分布型和中间分布型的复制比阈值分别采用“4.1.2”“4.1.3”和“4.1.4”中的推荐值，该流程易于掌握且实操性强。当然，各期刊社或编辑部也可根据自身特色，确定更加符合实际情况的不同类型论文复制比阈值，以多阈值替代目前的唯一阈值，使具体问题得到具体分析，让有价值的论文不被埋没，切实维护作者的发表权。

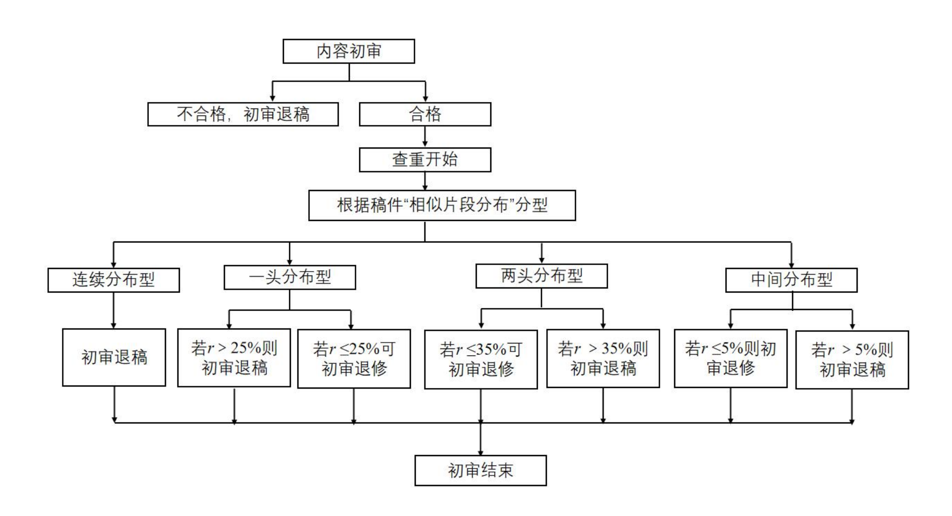


图 9　论文查重辩证取舍实操工作流程

在图 9 的基础上，画出基于辩证查重的计算机程序框图（见图 10），若根据图 10 编程并嵌入文献不端检测系统，可使论文查重工作趋向人工智能化。

作者简介：杨继涛(ORCID: 0000 － 0003 － 2896 － 8614)，硕士，副编审，《陕西农业科学》副主编，E-mail: snkx001@163.com
通讯作者郭柏寿(ORCID: 0000 － 0003 － 0119 － 2510)，副编审，《西北农业学报》副主编，E-mail: xbnx04@ nwsuaf. edu. cn

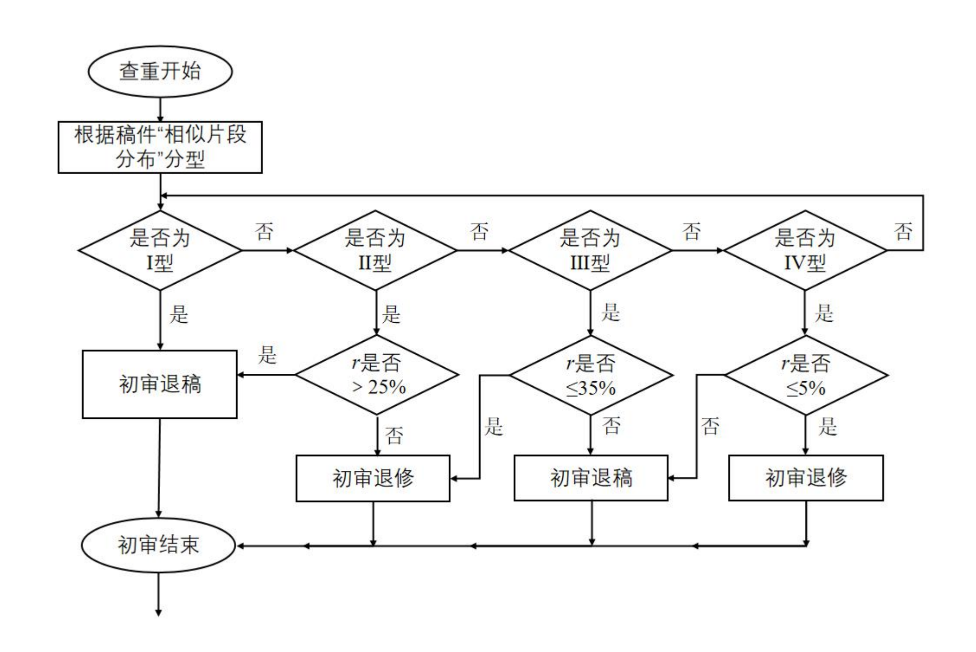


图 10　论文辩证查重取舍计算机程序框图

5 结语

初审查重不是小事，直接关乎论文是否能够进入同行评审以及复审与最后的终审环节，所以，初审工作中的粗枝大叶，看似无关紧要，但对作者的科研工作以及职业生涯的影响无疑是巨大的。因此，期刊编辑必须慎重对待每篇论文的初审查重。

在对论文总复制比解构分析的基础上，本文提出的查重辩证取舍实操工作流程，为期刊论文初审提供了参考，可使各期刊在论文初审环节的工作更加公平与公正，尽最大努力去保障作者的权益，不断提高期刊的服务水平。

笔者作为农业期刊编辑，查重复制比阈值设定范围的调查依托于 103 种综合性农业科技期刊，由于条件所限，没有将国内全部科技期刊作为调查对象进行研究。所以，对于不同类型论文复制比阈值的推荐值存在一定局限性，仅起到抛砖引玉的作用。不同的科技期刊可依据自身的学科分类、专业内容、行业特色等来确定科学合理的分型阈值。

论文审稿工作是期刊出版工作的重要环节之一，不断探索优化审稿过程的每个流程尤为必要。随着计算机和信息技术的不断发展，基于辩证分析思想的论文人工智能查重系统一定会被开发出来，长期困扰期刊的学术质量把控及学术不端行为界定等问题也会得到更加合理优化地解决，将使期刊论文的审稿环节更加合理、客观与公正。

参考文献：

[1]丛立先.相似度检测的学术、版权与社会意义[J].中国出版,2019(15):40-44.

[2]颜婧.论文重复率检测的法律规制研究[J].衡阳师范学院学报(社会科学),2021,42(1):85-92.

[3]孙娟,何 丽,宋勇刚,等.学术期刊在科研诚信建设中的作用与实施路径[J].中国科技期刊研究,2021,32(2)：153-157.

[4]田欣,马瀚青,郑军卫,等.国内外 5 种主要网络同行评议系统平台对比研究[J].中国科技期刊研

作者简介：杨继涛(ORCID: 0000 － 0003 － 2896 － 8614) ，硕士，副编审，《陕西农业科学》副主编，E-mail: snkx001@163.com
通讯作者郭柏寿(ORCID: 0000 － 0003 － 0119 － 2510) ，副编审，《西北农业学报》副主编，E-mail: xbnx04@ nwsuaf． edu． cn

究,2014,25(11):1363-1368.
[5]唐虹，朱银周.学术不端检测中文字复制比超限稿件的取舍分析[J].中国科技期刊研究,2020,31(3):281-287.
[6]曹梦园,王一涵,李男.科技期刊防范学术不端的分析与措施[J].标准实践,2022（1）（下）:92-94.
[7]李丹.学术不端检测系统的理性使用[J].学术界,2012(12):129-133,286.

Analyze Dialectically Duplicate Checking of Research Manuscripts, Ensure Justice and Equality of First Review

YANG Jitao[1)],GUO Baishou[2*)]
1） College of Economics & Management, Northwest A&F University, 3 Taicheng Ｒoad，Yangling 712100， China
2） ChinCollege of Agronomy, Northwest A&F University, 3 Taicheng Ｒoad，Yangling 712100，China

Abstract
[Objective]The purpose of this paper is to study the judgment methods for plagiarism detection o f scientific manuscripts in the first review, which is expected to accept or reject the manuscripts sc ientifically and dialectically, so as to make the first review of manuscripts more fair and just.**[Met hods]**Firstly, the survey statistical method is used to analyze the current situation of the first revie w, secondly, the total duplication ratio of the manuscripts is deconstructed by mathematical analys is, and finally the manuscripts are classified according to the distribution of similar fragments of m anuscripts, and different judgment methods are proposed according to different types. **[Result]**Th e survey and statistics show that manuscripts that exceed the specific threshold of total duplicatio n ratios will be rejected in the first review of journals ,which is biased. According to the importanc e of each part of the manuscripts and the distribution of similar fragments in different parts, the ma nuscripts can be divided into continuous distribution type, one-head distribution type, two-head di stribution type, and intermediate distribution type in the first review, among which the continuou s distribution type is divided into continuous uniform type and continuous non-uniform type. Th e one-head distribution type is further divided into the head distribution type and the tail distributi on type; On this basis, target judgment methods for different types are proposed, and the practica l workflow for plagiarism detection and computer block diagram for accepting or rejecting the ma nuscripts dialectically will be workout . **[Conclusion]** the first review of the manuscripts , the acad emic misconduct can not be judged by duplication ratio, duplication ratio should be analyzed diale ctically to protect the rights and interests of authors.
Key words Academic paper;Academic misconduct;Results of duplicate-checking;Justice and equality.

作者贡献声明：
1） 杨继涛：发现问题，查阅相关文献并整理分析，起草、完成论文初稿，修订论文 。
2） 郭柏寿：提出思路，并对论文进行最终把关修订。
作者简介：杨继涛(ＯＲCID: 0000 － 0003 － 2896 － 8614) ，硕士，副编审，《陕西农业科学》副主编，E-mail: snkx001@163.com
通讯作者郭柏寿(ＯＲCID: 0000 － 0003 － 0119 － 2510) ，副编审，《西北农业学报》副主编，E-mail: xbnx04@ nwsuaf． edu． cn

作者简介：杨继涛(ORCID: 0000 － 0003 － 2896 － 8614) ，硕士，副编审，《陕西农业科学》副主编，E-mail: snkx001@163.com
通讯作者郭柏寿(ORCID: 0000 － 0003 － 0119 － 2510) ，副编审，《西北农业学报》副主编，E-mail: xbnx04@ nwsuaf． edu． cn